Санкт-Петербургский государственный университет

Математика

Теория вероятностей и математической статистики

Шилов Илья Геннадьевич

# Критерии нормальности, основанные на свойствах нормального распределения

Выпускная квалификационная работа

Научный руководитель:

д. ф.-м. н., профессор Я. Ю. Никитин

Рецензент:

к. ф.-м. н., доцент В. В. Литвинова

Санкт-Петербург

2018

Saint Petersburg State University

Mathematics

Probability Theory and Mathematical Statistics

Shilov Ilia Gennadievich

# Normality tests based on a property of Normal Distribution

Graduation Project

Scientific Supervisor:

Professor D.Sci., Ya. Yu. Nikitin

Reviewer:

Associate Professor Ph.D., V. V. Litvinova

Saint Petersburg

2018

# Оглавление

## 1. Введение

Одной из важнейших областей математической статистики является проверка статистических гипотез. Хорошо известно, что нормальное распределение занимает центральное место в теории вероятностей и математической статистике, как в теоретических, так и практических задачах. Поэтому особенно важно иметь широкий набор инструментов для возможности проверки нормальности. Существует множество критериев проверки нормальности, основанных на самых разных идеях, см., например, [1], [2].

Одним из актуальных методов построения критериев нормальности является использование характеризаций, которых для нормального закона известно очень много, см., например, [3], [5].

Хорошо известным примером такой характеризации является знаменитая теорема Дж.Пойа [8]: пусть $X$ и $Y$ - центрированные независимые одинаково распределенные случайные величины; $X$ и $(X+Y)/\sqrt{2}$ имеют одинаковое распределение тогда и только тогда, когда $X$ и $Y$ нормально распределены с некоторой положительной дисперсией. Критерии нормальности, использующие эту характеризацию, были построены в [10], [12] и [11]. О других критериях нормальности, основанных на характеризациях, можно прочесть в [13]. Тем не менее, таких работ в настоящее время не так уж много, а область в целом можно назвать малоизученной. Исследования в этом направлении являются достаточно перспективными, а эффективности новых критериев зачастую довольно высоки относительно известных.

В данной работе мы строим статистический критерий согласия, основанный не на характеризации нормального закона, а на его известном свойстве, лишь близком к характеризации, которое будет описано ниже. В определенном смысле работа продолжает исследование Волковой и Никитина [14], в котором строились критерии экспоненциальности, основанные на свойстве экспоненциального закона, также не являющемся характеризацией. В работе [14] построенные критерии оказались весьма эффективными. Качество критериев, которые будут построены ниже для проверки нормальности, нам предстоит изучить.

Пусть $X_1, \dots, X_n$ - независимые наблюдения с нулевым средним и ф.р. $F$, а $F_n$ - эмпирическая функция распределения, построенная по этим наблюдениям. Нас интересует проверка сложной гипотезы $H_0 : F(x)$ - функция распределения нормального закона с нулевым средним и некоторой неизвестной дисперсией, против альтернативы $H_1$: $F$ - функция распределения закона, отличного от нормального.

Мы хотим использовать хорошо известное свойство нормальности: если $X, Y$ - независимые нормальные случайные величины с нулевым средним и некоторой дисперсией $\tau^2 > 0$, то случайная величина $X/Y$ имеет стандартное распределение Коши. Помимо этого, мы будем рассматривать величину $|X|/|Y|$, которая, соответственно, имеет одностороннее распределение Коши с функцией распределения $C(x) = \frac{2}{\pi}\arctan(x),\ x \geq 0$.

Известно, что это свойство - не характеризация нормальности. В 50-60-х годах была опубликована целая серия работ Молдона, Лаги и Котлярского, см., например, [15], [20], [21], в которых были построены примеры независимых, но негауссовских величин, отношение которых, тем не менее, имеет распределение Коши.

Однако множество таких законов, изучавшееся в указанных работах, по-видимому, не очень обширно. Поэтому критерии, построенные на указанном свойстве, вообще говоря, могут быть несостоятельными, но лишь для узкого набора специфических альтернатив. Но ведь многие критерии, хорошо известные в статистической практике, также несостоятельны против специфических альтернатив, например, критерий хи-квадрат, критерий Вилкоксона, критерий Джини и даже критерий отношения правдоподобия. Указанный недостаток становится также несущественным, если мы используем построенные критерии не для принятия, а для *отказа* от гипотезы нормальности.

Перед тем, как перейти к построению указанных критериев нормальности, основанных на $U$-эмпирических мерах, мы дадим краткое описание теории $U$-статистик.

## 2. Обзор теории $U$-статистик

Особым классом несмещенных оценок, играющих важную роль в математической статистике и теории вероятностей, являются $U$-статистики. Начало теории $U$-статистик восходит к фундаментальной работе В.Гефдинга [7], в которой вскрывается их структура и доказывается центральная предельная теорема для этого класса статистик.

В течение последующих лет интерес к этому классу постоянно возрастал и стало понятно, что множество важных и ценных статистик являются $U$-статистиками (или функционалами фон Мизеса, имеющими очень похожую асимптотическую теорию).

Мы рассматриваем $U$-статистики следующего вида:

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \dots < i_m \le n} \Psi(X_{i_1}, \dots, X_{i_m}), \qquad n \ge m,$$

где $X_1, X_2, \dots$ последовательность н.о.р.с.в. с функцией распределения $F$, в то время как ядро $\Psi : \mathbb{R}^m \to \mathbb{R}^1$ - измеримая функция от $m$ переменных. Без потери общности можно предполагать, что $\Psi$ является симметрической функцией относительно своих аргументов, так как в противном случае можно ввести симметрическое ядро $\Psi_0$ по формуле $\Psi_0(x_1, \dots, x_m) = \frac{1}{m!} \sum \Psi(x_{i_1}, \dots, x_{i_m})$, где суммирование осуществляется по всем $m!$ перестановкам чисел $(1, \dots, m)$. Число $m$ называется степенью ядра $\Psi$.

Предполагая ядро $\Psi$ интегрируемым на $\mathbb{R}^m$, определим регулярный функционал, заданный на пространстве вероятностных распределений:

$$\theta(F) := \int\limits_{\mathbb{R}^m} \!\cdots\! \int \Psi(x_1, \dots, x_m)\, dF(x_1) \dots dF(x_m). \tag{1}$$

Очевидно, что $U_n$ является несмещенной оценкой функционала $\theta(F)$, так как выполняется равенство $EU_n = \theta(F)$ для всех $F$ и $n \ge m$.

Приведем несколько примеров U-статистик:

1. Выборочное среднее: $\theta(F) = \mu(F) = \int x\, dF(x)$. Здесь $\Psi(x) = x, U_n = \frac{1}{n} \sum\limits_{j=1}^{n} X_j = \overline{X}$.

2. Выборочная дисперсия:

$$\theta(F) = \int (x - \mu)^2\, dF(x),$$

$$\Psi(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2 - 2x_1 x_2) = \frac{1}{2}(x_1 - x_2)^2$$

$$U_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \Psi(X_i, X_j) = \frac{1}{n-1} \left( \sum_{j=1}^{n} X_j{}^2 - n\overline{X} \right) = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X})^2.$$

3. Статистика Джини:

$$\theta(F) = E|X_1 - X_2|, \quad \Psi(x_1, x_2) = |x_1 - x_2|, \quad U_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} |X_i - X_j|.$$

Впоследствии нам понадобятся следующие обозначения:

$$\psi(x) := E_F(\Psi(X_1, \dots, X_m) | X_1 = x), \qquad \Delta^2 := E_F \psi^2(X_1) - (\theta(F))^2$$

где функция $\psi$ - одномерная проекция ядра $\Psi$, играющая важную роль в асимптотической теории $U$-статистик. Случай $\Delta^2 > 0$ соответствует невырожденности ядра. В этом случае мы можем использовать следующую знаменитую предельную теорему Гёффдинга [7]:

**Теорема 1.** *Пусть* $E_F\Psi^2(X_1,\dots,X_m) < \infty$ *и* $\Delta^2 > 0$. *Тогда при* $n \to \infty$ *имеет место следующая сходимость по распределению:*

$$\frac{\sqrt{n}}{m\Delta}(U_n - \theta(F)) \xrightarrow{d} N(0,1).$$

Если проекция равна нулю, то ядро $U$-статистики является вырожденным. Степень вырождения определяется так называемым рангом ядра. В случае вырожденных ядер предельное распределение $U$-статистик и функционалов Мизеса сильно усложняется, см. монографию [23].

В рассматриваемой задаче мы рассматриваем два типа статистик: интегральную статистику, асимптотически эквивалентную $U$-статистике с ядром степени 2, и статистику типа Колмогорова, которая является супремумом семейства $U$-статистик и устроена сложнее. Нас интересуют предельные распределения и большие уклонения обеих последовательностей статистик, а также их локальная бахадуровская эффективность при ряде альтернатив.

## 3. Эффективность по Бахадуру

Среди множества различных способов измерения асимптотической эффективности, таких как эффективность по Питмену, Ходжесу-Леману, Бахадуру, для наших целей лучше всего подходит эффективность по Бахадуру [24] в силу того, что, в отличие от питменовской эффективности, она может быть вычислена для статистик с предельным распределением, отличным от нормального, например, используемых в данной работе статистик типа Колмогорова. Эффективность Ходжеса-Лемана, в свою очередь, имеет другие недостатки, в частности, не различает двусторонние статистики и в целом является довольно грубой [25]. В этом разделе указаны краткие сведения о теории Бахадура.

Пусть $s = (X_1, X_2, \dots)$ - бесконечная последовательность независимых и одинаково распределенных случайных величин (н.о.р.с.в.) с распределением $P_\theta, \theta \in \Theta$ на $(\mathcal{X}, \mathcal{A})$. Пусть для проверки гипотезы

$$H_0 : \theta \in \Theta_0 \subset \Theta \subset \mathbb{R}^1$$

против альтернативы

$$H_0 : \theta \in \Theta_1 = \Theta \backslash \Theta_0$$

используется последовательность статистик $T_n(s) = T_n(X_1, \ldots, X_n)$. Согласно подходу Бахадура [24], следует зафиксировать мощность критериев и сравнить скорости экспоненциального убывания их размеров при увеличении объема выборки и фиксированной альтернативе. Характеристикой экспоненциальной скорости убывания для последовательности $\{T_n\}$ является некая детерминированная функция $c_T(\theta)$, $\theta \in \Theta_1$, называемая точным наклоном последовательности $\{T_n\}$.

Асимптотическая относительная эффективность по Бахадуру $e^B_{T,V}(\theta)$ последовательности $\{T_n\}$ по отношению к последовательности $\{V_n\}$ определяется по формуле

$$e^B_{T,V}(\theta) = c_T(\theta) \,/\, c_V(\theta). \tag{2}$$

Точный наклон, в свою очередь, может быть вычислен при помощи следующего фундаментального результата Бахадура:

**Теорема 2.** *Пусть выполняются следующие условия:*

1. $T_n \xrightarrow{P_\theta} b(\theta), \quad \theta > 0,$
   *где* $-\infty < b(\theta) < \infty$, *и* $\xrightarrow{P_\theta}$ *означает сходимость по вероятности;*

2. $\lim\limits_{n\to\infty} n^{-1} \ln \mathbb{P}_{H_0}(T_n \geq t) = -f(t)$
   *для любого* $\theta \in \Theta_0$ *и некоторого открытого интервала* $I$*, где функция* $f$ *непрерывна и* $\{b(\theta), \theta > 0\} \subset I$*.*

*Тогда при всех* $\theta \in \Theta_1$ *точный наклон* $c_T(\theta)$ *существует и вычисляется по формуле* $c_T(\theta) = 2f(b(\theta))$.

Часто не представляется возможным вычислить точную АОЭ по Бахадуру, однако мы можем вычислять локальную АОЭ при стремлении альтернативы к нулевой гипотезе, поскольку именно близкие гипотезы наиболее трудно и интересно различать. В этом случае мы говорим о локальной асимптотической эффективности Бахадура (ЛАЭ) и о локальных точных наклонах.

Важную роль при проверке статистических гипотез играет расстояние Кульбака-Лейблера $K(\theta, \Theta_0)$ между истинным и альтернативным распределениями. Аналогом неравенства Крамера-Рао в нашем случае выступает неравенство Бахадура-Рагавачари, дающее верхнюю оценку точного наклона:

$$c_T(\theta) \leq \inf\{2K(\theta, \Theta_0),\ \theta_0 \in \Theta_0\}, \quad \forall \theta \in \Theta_1.$$

Следовательно, мы можем определить локальную эффективность Бахадура по формуле:

$$e_T = \lim_{\theta \to \partial\Theta_0} c_T(\theta) \,/\, 2K(\theta, \Theta_0). \tag{3}$$

Интересным представляется случай, когда $c_T(\theta) \sim 2K(\theta, \Theta_0), \theta \to \partial\Theta_0$, и, соответственно,

$$e_T = \lim_{\theta \to \partial\Theta_0} c_T(\theta) \,/\, 2K(\theta, \Theta_0) = 1.$$

В этом случае можно говорить о локальной асимптотической оптимальности по Бахадуру. Описанию альтернатив, при которых рассматриваемый критерий локально оптимален, мы уделим внимание ниже.

Первое условие Теоремы 2 - разновидность закона больших чисел и проверяется довольно легко. В то же время второе условие теоремы - грубая асимптотика вероятностей больших уклонений, и её вычисление может представлять большие сложности. Мы будем использовать для анализа вводимых ниже статистик следующую теорему [26]:

**Теорема 3.** *Пусть $V_n$ - последовательность U-статистик с невырожденным, ограниченным и центрированным ядром $\Psi$. Тогда*

$$\lim_{n \to \infty} n^{-1} \ln \mathbb{P}(V_n \geq a) = -\sum_{j=2}^{\infty} b_j a^j \tag{4}$$

*где ряд в правой части сходится при достаточно малых $a > 0$, а $b_2 = (2m^2\Delta^2)^{-1}$, где $\Delta^2$ - дисперсия проекции ядра $\Psi$.*

## 4. Интегральная статистика $W_n^1$

Рассмотрим вместо стандартной эмпирической функции распределения

$$F_n(t) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i < t\}, t \geq 0,$$

$U$ - эмпирическую функцию распределения

$$H_n(t) = \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} \left( \mathbf{1} \left\{ \frac{X_i}{X_j} < t \right\} + \mathbf{1} \left\{ \frac{X_j}{X_i} < t \right\} \right).$$

Известно, что свойства $U$ - эмпирической функции распределения во многом аналогичны свойствам стандартной эмпирической ф.р. Следовательно для больших $n$ при условии $H_0$ ф.р. $H_n$ должна быть близка к закону Коши. Рассмотрим интегральную статистику $W_n^1$

$$W_n^1 = \int_{-\infty}^{\infty} (C(t) - H_n(t)) q(t) \, dt$$

где $q(t)$ - некоторая весовая функция, необходимая для сходимости интеграла. В качестве веса в данной работе выбирается функция $q(t) = |t| e^{-t^2/2}$. Первоначально мы рассматривали вес

$$q(t) = \frac{|t|}{(1+t^2)^3},$$

удобный для вычислений интегралов с плотностью Коши. Однако в процессе работы выяснилось, что скорость убывания этого веса $O(t^{-5})$ на бесконечности недостаточна для сходимости всех появляющихся интегралов. Поэтому было решено перейти к другому, быстро убывающему весу, а именно к $q(t) = |t| e^{-t^2/2}$.

Тогда вычисления ядра $U$-статистики примут следующий вид:

$$
\begin{aligned}
W_n^1 &= \int\limits_{-\infty}^{\infty} (C(t) - H_n(t))q(t)\,dt \\
&= \int\limits_{-\infty}^{\infty} \left(\frac{1}{\pi}\arctan(t) + \frac{1}{2}\right) q(t)\,dt - \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \int\limits_{-\infty}^{\infty} \left(\mathbf{1}\left\{\frac{X_i}{X_j} < t\right\} + \mathbf{1}\left\{\frac{X_j}{X_i} < t\right\}\right) q(t)\,dt \\
&= \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \left(1 - \frac{1}{2}\left(\int\limits_{\frac{X_i}{X_j}}^{0} \mathbf{1}\left\{\frac{X_i}{X_j} < 0\right\}(-t)e^{-t^2/2}\,dt + \int\limits_{0}^{\infty} \mathbf{1}\left\{\frac{X_i}{X_j} < 0\right\} te^{-t^2/2}\,dt\right.\right. \\
&+ \int\limits_{\frac{X_i}{X_j}}^{\infty} \mathbf{1}\left\{\frac{X_i}{X_j} > 0\right\} q(t)\,dt + \int\limits_{\frac{X_j}{X_i}}^{0} \mathbf{1}\left\{\frac{X_j}{X_i} < 0\right\}(-t)e^{-t^2/2}\,dt + \int\limits_{0}^{\infty} \mathbf{1}\left\{\frac{X_j}{X_i} < 0\right\} te^{-t^2/2}\,dt \\
&\left.+ \int\limits_{\frac{X_j}{X_i}}^{\infty} \mathbf{1}\left\{\frac{X_j}{X_i} > 0\right\} q(t)\,dt\right) \\
&= \frac{2}{n(n-1)} \sum_{1\le i<j\le n} 1 - \frac{1}{2}\left(2\left(\mathbf{1}\left\{\frac{X_i}{X_j} < 0\right\} + \mathbf{1}\left\{\frac{X_j}{X_i} < 0\right\}\right) - \mathbf{1}\left\{\frac{X_i}{X_j} < 0\right\} e^{-\frac{X_i^2}{2X_j^2}}\right. \\
&\left.\left. - \mathbf{1}\left\{\frac{X_j}{X_i} < 0\right\} e^{-\frac{X_j^2}{2X_i^2}} + \mathbf{1}\left\{\frac{X_i}{X_j} > 0\right\} e^{-\frac{X_i^2}{2X_j^2}} + \mathbf{1}\left\{\frac{X_j}{X_i} > 0\right\} e^{-\frac{X_j^2}{2X_i^2}}\right)\right).
\end{aligned}
$$

Следовательно, мы получили $U$ - статистику с ядром степени 2:

$$
\begin{aligned}
\Psi(X,Y) = 1 - \frac{1}{2}\Bigg(2\left(\mathbf{1}\left\{\frac{X}{Y} < 0\right\} + \mathbf{1}\left\{\frac{Y}{X} < 0\right\}\right) - \mathbf{1}\left\{\frac{X}{Y} < 0\right\} e^{-\frac{X^2}{2Y^2}} \\
- \mathbf{1}\left\{\frac{Y}{X} < 0\right\} e^{-\frac{Y^2}{2X^2}} + \mathbf{1}\left\{\frac{X}{Y} > 0\right\} e^{-\frac{X^2}{2Y^2}} + \mathbf{1}\left\{\frac{Y}{X} > 0\right\} e^{-\frac{Y^2}{2X^2}}\Bigg).
\end{aligned}
$$

Учитывая тождество $X/Y < 0 \Leftrightarrow Y/X < 0$ перепишем ядро так:

$$
\begin{aligned}
\Psi(X,Y) = \frac{e^{-\frac{X^2}{2Y^2}}}{2}\left(\mathbf{1}\left\{\frac{X}{Y} < 0\right\} - \mathbf{1}\left\{\frac{X}{Y} > 0\right\}\right) + \frac{e^{-\frac{Y^2}{2X^2}}}{2}\left(\mathbf{1}\left\{\frac{Y}{X} < 0\right\} - \mathbf{1}\left\{\frac{Y}{X} > 0\right\}\right) \\
+ 1 - 2\cdot\mathbf{1}\left\{\frac{X}{Y} < 0\right\}.
\end{aligned}
\tag{5}
$$

При вычислении проекции ядра $\psi(s) := E(\Psi(X,Y)|Y = s)$, учитывая, что математическое ожидание индикатора $\mathbf{1}(A)$ равно $\mathbb{P}(A)$, и в силу симметричности распределения нормального закона, получаем, вне зависимости от знака $s$:

$$
\psi(s) = 0. \tag{6}
$$

Таким образом наша $U$-статистика оказывается вырожденной. В этом случае предельное распределение и большие уклонения определяются собственными числами интегрального оператора Фредгольма с ядром (5), которые нам неизвестны и не видно, как их найти. Таким образом, этот путь в настоящее время не приводит к успеху.

## 5. Интегральная статистика $W_n^2$

Рассматривая отношение нормальных величин, с учетом предыдущего раздела, разумно рассмотреть отношение их модулей. По существу, проверка нормальности выборки эквивалентна проверке "правосторонней нормальности" абсолютных значений той же выборки.

Поэтому вместо стандартной эмпирической функции распределения рассмотрим $U$-эмпирическую функцию распределения, построенную на модулях нормальных величин:

$$\hat{H}_n(t) = \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \left( \mathbf{1}\left\{ \frac{|X_i|}{|X_j|} < t \right\} + \mathbf{1}\left\{ \frac{|X_j|}{|X_i|} < t \right\} \right).$$

Для больших n при условии $H_0$ ф.р. $\hat{H}_n$ должна быть близка к одностороннему закону Коши с ф.р. $\frac{2}{\pi}\arctan(t), t \ge 0$.

В качестве веса выберем $q(t) = te^{-t^2/2}, t \ge 0$, соответственно рассмотрим интегральную статистику

$$\begin{aligned}
W_n^2 &= \int_0^\infty (C(t) - H_n(t))q(t)\,dt \\
&= \int_0^\infty C(t)q(t)\,dt + \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \left( \int_{\frac{|X_i|}{|X_j|}}^\infty dQ(t) + \int_{\frac{|X_j|}{|X_i|}}^\infty dQ(t) \right) \\
&= \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \left( \sqrt{e}\,\mathrm{Erfc}\,\frac{1}{\sqrt{2}} - \frac{1}{2} e^{-\frac{X_i^2}{2X_j^2}} - \frac{1}{2} e^{-\frac{X_j^2}{2X_i^2}} \right).
\end{aligned}$$

Мы воспользовались следующими вычислениями:

$$\begin{aligned}
\int_0^\infty C(t)q(t)\,dt &= \frac{2}{\pi}\int_0^\infty t\,\mathrm{arctg}\,t\,e^{-\frac{t^2}{2}}\,dt = -\frac{2}{\pi}\int_0^\infty \mathrm{arctg}\,t\,de^{-\frac{t^2}{2}} \\
&= -\frac{2}{\pi}\,\mathrm{arctg}\,t\,e^{-\frac{t^2}{2}}\Big|_0^\infty + \frac{2}{\pi}\int_0^\infty \frac{1}{(1+t^2)} e^{-\frac{t^2}{2}}\,dt.
\end{aligned}$$

Последний интеграл может быть найден в [18]; следовательно, мы получили $U$-статистику с ядром степени 2:

$$\Psi(|X|,|Y|) = \sqrt{e}\,\mathrm{Erfc}\,\frac{1}{\sqrt{2}} - \frac{1}{2}e^{-\frac{X^2}{2Y^2}} - \frac{1}{2}e^{-\frac{Y^2}{2X^2}}. \tag{7}$$

С учетом обозначений $\hat{X} := |X|, \hat{Y} := |Y|$ проекция ядра имеет вид:

$$\psi(s) := E(\Psi(\hat{X},\hat{Y})|\hat{Y} = s) = \sqrt{e}\,\mathrm{Erfc}\,\frac{1}{\sqrt{2}} - \frac{1}{2}Ee^{-\frac{\hat{X}^2}{2s^2}} - \frac{1}{2}Ee^{-\frac{s^2}{2\hat{X}^2}}.$$

В то время, как первое ожидание вычисляется довольно просто, для вычисления второго нам понадобится результат из [6]:

$$\int\limits_0^\infty \exp(-ax^2 - \frac{b}{x^2})\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\exp(-2\sqrt{ab}),$$

где в нашем случае параметры $a, b$ равны соответственно $\frac{1}{2}$ и $\frac{s^2}{2}$. Таким образом, вычисления проекции дают следующий результат:

$$\psi(s) := \sqrt{e}\,\mathrm{Erfc}\,\frac{1}{\sqrt{2}} - \frac{1}{2}\frac{|s|}{\sqrt{s^2+1}} - \frac{1}{2}e^{-|s|}. \tag{8}$$

Среднее проекции $E\psi(X) = 0$, как и должно быть.

Дисперсия проекции может быть вычислена с помощью численных методов по формуле ниже, где $\overline{\varphi}(s) = 2\varphi(s), s \geq 0$ - плотность полунормального распределения:

$$\Delta^2 = \int\limits_0^\infty \psi(s)^2\overline{\varphi}(s)\,ds.$$

В нашем случае $\Delta^2 = 0.00015965...$. Таким образом, ядро $\Psi$ центрировано и невырождено; применяя Теорему 1, получаем следующий результат:

**Теорема 4.** *При гипотезе $H_0$ и $n \to \infty$, статистика $\sqrt{n}W_n$ асимптотически нормальна*

$$\sqrt{n}W_n \xrightarrow{d} \mathcal{N}(0, 4\Delta_W^2)$$

*где* $\Delta_W^2 = 0.00015965$.

Аналогично мы можем применить Теорему 3, и таким образом получим асимптотику больших уклонений, необходимую для выполнения условия (2) Теоремы 2:

$$\lim_{n\to\infty} n^{-1}\ln \mathbb{P}_{H_0}(T_n \geq a) = -f(a),$$

где

$$f(a) = \frac{a^2}{8\Delta^2}(1 + o(1)) \sim 782.962a^2,\ a \to 0.$$

Для вычисления точного наклона осталось проверить выполнение условия (1). По теореме Гливенко-Кантелли для $U$-статистик:

$$\begin{aligned} b_W(\theta) &= \int_0^\infty (C(t) - F(t))\, q(t)\, dt = \int_0^\infty \left( C(t) - \mathbb{P}_\theta \left( \frac{|X|}{|Y|} < t \right) \right) q(t)\, dt \\ &= \int_0^\infty \left( \frac{2}{\pi} \arctan t - \int_{-t}^{t} f(z, \theta)\, dz \right) q(t)\, dt, \end{aligned} \tag{9}$$

где $f(z, \theta)$ - плотность распределения $\frac{X}{Y}$ при альтернативе. Рассмотрим альтернативу сдвига, при которой с.в. имеют ф.р. $\Phi(x - \theta)$. Тогда согласно [9] плотность $\frac{X}{Y}$ имеет следующий вид:

$$f(z) = \frac{e^{-\theta^2}}{\pi(1 + z^2)} \left( 1 + qe^q \int_0^q e^{-x^2/2}\ dx \right), \quad \text{где} \quad q = \frac{\theta(z + 1)}{\sqrt{1 + z^2}}. \tag{10}$$

Раскладывая $f(z)$ в ряд по $\theta$ в окрестности 0, получаем

$$\int_0^\infty \left( \frac{2}{\pi} \arctan t - \int_{-t}^{t} \left( \frac{1}{\pi(1 + z^2)} + \frac{2z\theta^2}{\pi(1 + z^2)^2} + O(\theta^3) \right)\ dz \right) q(t)\, dt.$$

Как видно из формулы, слагаемое, содержащее $\theta$ в этом случае равно 0. Значит, $b(\theta)$ имеет порядок $O(\theta^2)$, а точный наклон имеет порядок $O(\theta^4)$.

Поскольку расстояние Кульбака-Лейблера между гипотезой и альтернативой сдвига имеет для нормального закона порядок $O(\theta^2)$, то таким образом, эффективность критерия при данной альтернативе равна 0.

Ниже рассмотрены некоторые типичные для нормального распределения альтернативы и, как можно увидеть, слагаемые при $\theta$ в разложении также равны 0, и, следовательно, эффективность критерия снова нулевая.

Рассмотрим скошенную альтернативу по Аззалини [16] $2\varphi(x)\Phi(\theta x)$:

$$\begin{aligned} f_{X/Y}(z) &= \int_{-\infty}^{\infty} 4|x|\varphi(x)\Phi(\theta x)\varphi(zx)\Phi(\theta zx)\, dx \\ &= \int_{-\infty}^{\infty} \left( \frac{e^{-x^2(z^2+1)/2}|x|}{2\pi} + \frac{e^{-x^2(z^2+1)/2}x(1 + z)|x|\theta}{\pi\sqrt{2\pi}} + \frac{e^{-x^2(z^2+1)/2}x^2 z|x|\theta^2}{\pi^2} + O(\theta^3) \right) dx \\ &= \frac{1}{\pi(1 + z^2)} + 0 + \frac{4\theta^2 z}{\pi^2(1 + z^2)^2}. \end{aligned}$$

Аналогично для альтернативы загрязнения, при которой функция распределения имеет вид $(1-\theta)\Phi(x)+\theta\Phi^2(x)$:

$$
\begin{aligned}
f_{X/Y}(z) &= \int\limits_{-\infty}^{\infty} |x|((1-\theta)\varphi(x)+2\theta\varphi(x)\Phi(x))((1-\theta)\varphi(zx)+2\theta\varphi(zx)\Phi(zx))\,dx \\
&= \int\limits_{-\infty}^{\infty} \left(\frac{e^{-x^2(z^2+1)/2}|x|}{2\pi} + \theta\frac{e^{-x^2(z^2+1)/2}|x|(Erf(\frac{x}{\sqrt{2}})+Erf(\frac{zx}{\sqrt{2}}))}{2\pi} + O(\theta^2)\right) dx \\
&= \frac{1}{\pi(1+z^2)}.
\end{aligned}
$$

Рассматривая в качестве альтернативной плотности предложенную в [17] асимметричную плотность следующего вида

$$
f(x,\theta) = \varphi(\frac{x}{1+\theta})I(x<0) + \varphi(\frac{x}{1-\theta})I(x\geq 0),
$$

получаем аналогичный результат:

$$
\begin{aligned}
f_{Z=X/Y}(z) &= \int\limits_{0}^{\infty} x\varphi(\frac{x}{1-\theta})\varphi(\frac{zx}{1-\theta})\,dx - \int\limits_{-\infty}^{0} x\varphi(\frac{x}{1+\theta})\varphi(\frac{zx}{1+\theta})\,dx \\
&= \frac{(\theta-1)^2}{2\pi(1+z^2)} - \left(-\frac{(\theta+1)^2}{2\pi(1+z^2)}\right) \\
&= \frac{\theta^2+1}{\pi(1+z^2)}
\end{aligned}
$$

Предполагая, что нулевая эффективность получается из-за излишней симметричности альтернатив, рассмотрим несколько испорченную плотность вида

$$
f(x,\theta) = (1+\theta)\varphi(\frac{x}{1+\theta})I(x<0) + \frac{1-2\theta-\theta^2}{1-\theta}\varphi(\frac{x}{1-\theta})I(x\geq 0). \tag{11}
$$

Тем не менее, получаем точно такой же результат:

$$
\begin{aligned}
f_{Z=X/Y}(z) &= \int\limits_{0}^{\infty} x\varphi(\frac{x}{1-\theta})\varphi(\frac{zx}{1-\theta})\left(\frac{1-2\theta-\theta^2}{1-\theta}\right)^2 dx - \int\limits_{-\infty}^{0} (1+\theta)^2 x\varphi(\frac{x}{1+\theta})\varphi(\frac{zx}{1+\theta})\,dx \\
&= \frac{(1-2\theta-\theta^2)^2}{\pi(1+z^2)} - \left(-\frac{(1+\theta)^4}{2\pi(1+z^2)}\right) \\
&= \frac{1+4\theta^2+4\theta^3+\theta^4}{\pi(1+z^2)}.
\end{aligned}
$$

Для многих альтернатив, удовлетворяющих условиям регулярности, полезным инструментом для нахождения $b(\theta)$ является следующий результат.

**Лемма 1.** *При двукратной гладкости по параметру $\theta$ ф.р. при альтернативе, где $\theta \in \Theta = [0, \theta^*]$ верно следующее:*

$$b(\theta) = 2\theta \int\limits_0^\infty \psi(x) f_\theta(x; 0)\, dx. \tag{12}$$

*где $f_\theta$ - производная плотности при альтернативе.*

*Доказательство.* Согласно закону больших чисел для U-статистик [23], $U_n \to \theta(F)$, где $\theta(F)$ в нашем случае, согласно (1):

$$\theta(F) = \int\limits_0^\infty \int\limits_0^\infty \Psi(x, y) f(x; \theta) f(y; \theta)\, dx\, dy.$$

Введя обозначение $L(x, y, \theta) = f(x; \theta) f(y; \theta)$ и пользуясь разложением функции в ряд Тейлора с остатком в интегральной форме по $\theta$ при $\theta \to 0$, по аналогии с [22] получаем следующий результат:

$$L(x, y, \theta) = L(x; 0) + \theta L_\theta(x; 0) + \frac{1}{2}\theta^2 \int\limits_0^1 t^2 (1-t)^2 L_{\theta\theta}(x; t\theta)\, dt,$$

где

$$L(x, y, 0) = f(x; 0) f(y; 0), \qquad L_\theta(x, y, 0) = f(x; 0) f_\theta(y; 0) + f(y; 0) f_\theta(x; 0).$$

Формула для $L_{\theta\theta}(x; z)$, где $y \in [0, \theta^*)$:

$$L_{\theta\theta}(x, y, z) = f_{\theta\theta}(x; z) f(y, z) + f_{\theta\theta}(y; z) f(x, z) + 2 f_\theta(y; z) f_\theta(x; z).$$

Таким образом, применяя теорему Фубини, получаем следующий результат:

$$b(\theta) = 2\theta \int\limits_0^\infty \psi(x) f_\theta(x; 0)\, dx,$$

что завершает доказательство леммы. □

Может показаться, что разработанный нами интегральный критерий совершенно бесполезен, поскольку для многих ходовых альтернатив его асимптотическая эффективность равна нулю. Однако это не так, и для специальных альтернатив эта эффективность может оказаться высокой и даже максимально возможной.

Рассмотрим теперь случай локальной асимптотической оптимальности по Бахадуру, при котором эффективность критерия равна 1. Для этого введем специальную альтернативу вида

$$f(x, \theta) = \overline{\varphi}(x) + \theta \psi(x) \overline{\varphi}(x), \theta \geq 0,\ x \geq 0.$$

При малых $\theta$ это, очевидно, неотрицательная функция, а в силу того, что $E\psi(X) = 0$, интеграл от нее равен 1, так что это действительно плотность.

Далее, $f_\theta(x, 0) = \psi(x)\overline{\varphi}(x)$, но тогда

$$b(\theta) \sim 2\theta \int\limits_0^\infty \psi^2(x)\overline{\varphi}(x)\, dx = 2\theta\Delta^2.$$

Отсюда уже, в силу формулы (6) работы [22], локальный точный наклон нашей статистики будет равен

$$c_W(\theta) \sim \Delta^2\theta^2,\ \theta \to 0.$$

Остается просчитать расстояние Кульбака-Лейблера $K(f(x;\theta), \overline{\varphi}(x))$ для нашей специальной альтернативы, которое будет иметь следующий вид:

$$K(\theta) = \int\limits_0^\infty \ln(1 + \theta\psi(x))(\overline{\varphi}(x) + \theta\psi(x)\overline{\varphi}(x))\, dx \sim \frac{1}{2}\theta^2 \int\limits_0^\infty \psi^2\overline{\varphi}(x)\, dx,$$

где последний интеграл как раз равен $\Delta^2$. По формуле (3)

$$e_W = c_W(\theta)/2K(\theta) = 1,$$

что и дает локальную асимптотическую эффективность по Бахадуру.

## 6. Статистика типа Колмогорова $D_n^2$

По аналогии с интегральной статистикой $W_n^2$, рассмотрим теперь статистику типа Колмогорова $D_n^2$, определенную следующим образом:

$$D_n^2 = \sup_{t \geq 0} \left| C(t) - \hat{H}_n(t) \right| \tag{13}$$

При фиксированном $t$ разность $C(t) - \hat{H}_n(t)$ является семейством U-статистик с ядром $\Xi$, зависящим от $t \geq 0$:

$$\Xi(X_i, X_j, t) = \frac{2}{\pi} \arctan t - \frac{1}{2} \mathbf{1}\left\{ \frac{|X_i|}{|X_j|} < t \right\} - \frac{1}{2} \mathbf{1}\left\{ \frac{|X_j|}{|X_i|} < t \right\}. \tag{14}$$

С учетом обозначений $\hat{X} := |X|, \hat{Y} := |Y|$ проекция ядра имеет следующий вид:

$$\xi(s;t) = E(\Xi(\hat{X}, \hat{Y}, t) | \hat{X} = s) = \frac{2}{\pi} \arctan t - \frac{1}{2} \mathbb{P}\left( \frac{|s|}{\hat{Y}} < t \right) - \frac{1}{2} \mathbb{P}\left( \frac{\hat{Y}}{|s|} < t \right)$$

После несложных вычислений получаем выражение для семейства проекций:

$$\xi(s;t) = \frac{2}{\pi} \arctan t + \Phi\left( \frac{s}{t} \right) - \Phi(st) - \frac{1}{2}, \tag{15}$$

где $\Phi(x)$ означает функцию распределения нормального закона. При вычислении среднего проекции, воспользуемся результатами [19]:

$$\begin{aligned} \int_0^\infty \Phi(\frac{s}{t}) \overline{\varphi}(s)\, ds &= \frac{1}{2} + \frac{\arctan 1/t}{\pi} \\ \int_0^\infty \Phi(st) \overline{\varphi}(s)\, ds &= \frac{1}{2} + \frac{\arctan t}{\pi}. \end{aligned} \tag{16}$$

Благодаря хорошо известному равенству $\operatorname{arctg} x + \operatorname{arctg} \frac{1}{x} = \frac{\pi}{2} sign(x)$ мы можем вычислить среднее проекции (15), равное в данном случае нулю. Для вычисления дисперсии $\delta^2(t) = E\xi^2(X;t)$ при гипотезе $H_0$ нам понадобится следующая лемма:

**Лемма 2.**

$$\int_0^\infty \Phi(\frac{s}{t}) \Phi(st)\, d\overline{\varphi}(s) = \frac{1}{2} + \frac{1}{2\pi} \operatorname{arctg} \frac{t}{\sqrt{t^4 + t^2 + 1}}.$$

*Доказательство.* Введём следующее обозначение:

$$I(t) = \int_0^\infty \Phi(\frac{s}{t}) \Phi(st)\, d\overline{\varphi}(s),$$

где $\Phi(x)$ - функция распределения нормального закона. Дифференцируя $I(t)$ по $t$, имеем

$$I'(t) = I_1(t) - I_2(t),$$

где

$$I_1(t) = C(\pi)\int\limits_0^\infty s\varphi(s\sqrt{1+t^2})\Phi(\frac{s}{t})\,ds = -\frac{C(\pi)}{(1+t^2)}\int\limits_0^\infty \Phi(\frac{s}{t})\,d\varphi(s\sqrt{t^2+1}) =$$

$$= -\frac{C(\pi)}{(1+t^2)}\left(-\frac{1}{2} - \frac{1}{t}\int\limits_0^\infty \varphi(s\sqrt{t^2+1})\varphi(\frac{s}{t})\,ds\right) = \frac{C(\pi)}{2(1+t^2)}\left(1+\frac{1}{\sqrt{t^4+t^2+1}}\right).$$

Аналогично

$$I_2(t) = \frac{C(\pi)}{t^2}\int\limits_0^\infty s\varphi(s\frac{\sqrt{t^2+1}}{t})\Phi(st)\,ds = \frac{C(\pi)}{2(1+t^2)}\left(1+\frac{t^2}{\sqrt{t^4+t^2+1}}\right).$$

Таким образом, мы получаем

$$I'(t) = C(\pi)\frac{1}{t^2+1}\left(\frac{1-t^2}{\sqrt{t^4+t^2+1}}\right),$$

где $C(\pi) = \frac{1}{2\pi}$. Решая дифференциальное уравнение $I(t) + C = \int I'(t)\,dt$ мы получаем следующий результат:

$$I(t) = \frac{1}{2} + \frac{1}{2\pi}\operatorname{arctg}\frac{t}{\sqrt{t^4+t^2+1}}, \tag{17}$$

что завершает доказательство леммы. $\square$

Согласно [19]:

$$\begin{aligned} \int\limits_0^\infty \Phi^2(\frac{s}{t})\,d\overline{\varphi}(s) &= \frac{1}{\pi}\left(\operatorname{arctg}\frac{1}{t} + \operatorname{arctg}\sqrt{1+\frac{2}{t^2}}\right), \\ \int\limits_0^\infty \Phi^2(st)\,d\overline{\varphi}(s) &= \frac{1}{\pi}\left(\operatorname{arctg} t + \operatorname{arctg}\sqrt{1+2t^2}\right). \end{aligned} \tag{18}$$

Таким образом, используя формулы (16), (17), (18) получаем выражение для $\delta^2(t)$:

$$\begin{aligned} \delta^2(t) = &\frac{4}{\pi^2}\operatorname{arctg}^2 t + \frac{1}{4} + \frac{1}{\pi}\left(\operatorname{arctg}\frac{1}{t} + \operatorname{arctg}\sqrt{1+\frac{2}{t^2}}\right) \\ &+ \frac{1}{\pi}\left(\operatorname{arctg} t + \operatorname{arctg}\sqrt{1+2t^2}\right) + \frac{4}{\pi}\operatorname{arctg} t\left(\frac{1}{2} + \frac{\operatorname{arctg}\frac{1}{t}}{\pi}\right) \\ &- \frac{4}{\pi}\operatorname{arctg} t\left(\frac{1}{2} + \frac{\operatorname{arctg} t}{\pi}\right) - \left(\frac{1}{2} + \frac{\operatorname{arctg}\frac{1}{t}}{\pi}\right) + \left(\frac{1}{2} + \frac{\operatorname{arctg} t}{\pi}\right) \\ &- \frac{2\operatorname{arctg} t}{\pi} - 2I(t) \\ &= \frac{1}{4} + \frac{1}{\pi}\operatorname{arctg}\sqrt{1+\frac{2}{t^2}} + \frac{4}{\pi^2}\operatorname{arctg} t\operatorname{arctg} 1/t + \frac{1}{\pi}\operatorname{arctg}\sqrt{1+2t^2} - 2I(t). \end{aligned}$$

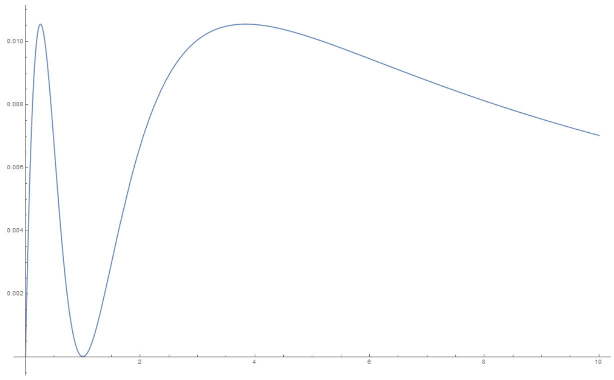

Рис. 1. Plot of $\delta^2$

Используя пакет WOLFRAM, получаем следующие значения для нулей $(\delta^2(t))'$ и для максимума $\delta^2(t)$:

$$\delta^2(t)' = 0 \quad \text{при} \quad t = 1,\, t = 0.2605,\, t = 3.83878.$$

$$\sup_{t \geq 0} \delta^2(t) = 0.0105485.$$

Применяя аналог теоремы о больших уклонениях для статистик типа Колмогорова [27], получаем следующий результат:

**Теорема 5.** *Для $a > 0$ при гипотезе $H_0$*

$$\lim_{n\to\infty} n^{-1} \ln \mathsf{P}(D_n > a) = -f_D(a),$$

*где функция $f_D$ непрерывна для достаточно малого $a > 0$; также*

$$f_D(a) = \frac{a^2}{8\delta^2}(1 + o(1)) \sim 11.85a^2 \text{ при } \quad a \to \ 0.$$

Для вычисления предела $b_D(\theta)$ сначала вычислим предел при альтернативе $b_D(t, \theta)$ под знаком супремума, для чего воспользуемся следующей леммой:

**Лемма 3.** *Для альтернативы с плотностью распределения $g(x, \theta)$, имеем:*

$$b_D(t, \theta) \sim 2\theta \int_0^\infty \xi(x; t) h(x)\, dx, \quad h(x) = g'_\theta(x, 0).$$

*Доказательство.* По теореме Гливенко-Кантелли для $U$-статистик предел по вероятности выражения под супремумом статистики $D_n$ при альтернативе $H_1$ равен

$$b_D(t, \theta) = \frac{2}{\pi} \operatorname{arctg} t - \mathsf{P}_\theta \left( \frac{\hat{X}}{\hat{Y}} < t \right).$$

Тогда при $\theta \to 0$

$$b_D(t,\theta) = \frac{2}{\pi}\operatorname{arctg} t - \int\limits_0^\infty g(y,\theta)\,dy \int\limits_0^{yt} g(x,\theta)\,dx \sim b_D(t,0) + b'_D(t,0)\theta.$$

Легко увидеть, что

$$b_D(t,0) = 0,$$
$$b'_D(t,0) = -\int\limits_0^\infty h(y)\,dy \int\limits_0^{yt} \overline{\varphi}(x)\,dx - \int\limits_0^\infty \overline{\varphi}(y)\,dy \int\limits_0^{yt} h(x)\,dx.$$

Меняя порядок интегрирования во втором интеграле, получаем:

$$\begin{aligned} b'_D(t,0) &= -\int\limits_0^\infty (2\Phi(yt)-1)h(y)\,dy - \int\limits_0^\infty h(x)\,dx \int\limits_{x/t}^\infty \overline{\varphi}(y)\,dy \\ &= -2\int\limits_0^\infty h(y)(\Phi(yt) + 1 - \Phi(y/t))\,dy \\ &= 2\int\limits_0^\infty \xi(x;t)h(x)\,dx \end{aligned}$$

Таким образом

$$b_D(\theta) := \sup_{t\ge 0} |b_D(t,\theta)| \sim \sup_{t\ge 0}\left|2\theta\int\limits_0^\infty \xi(x;t)h(x)\,dx\right|,$$

что завершает доказательство леммы. □

Очевидно, что для альтернатив, указанных в разделе про интегральную статистику $W_n^2$, эффективность аналогично будет равна 0.

По аналогии с интегральной статистикой $W_n{}^2$, рассмотрим случай локальной асимптотической эффективности по Бахадуру. Для этого в качестве альтернативной плотности возьмем

$$g(x;\theta) = \overline{\varphi}(x) + \theta\xi(x;t_0)\overline{\varphi}(x), \quad \theta \ge 0, x \ge 0, \tag{19}$$

где $t_0 = \mathrm{argmax}_{t\ge 0}\delta^2(t)$. В этом случае расстояние Кульбака-Лейблера будет равно

$$K(\theta) = \sup_{t\ge 0}\int\limits_0^\infty \ln(1+\theta\xi(x;t_0))(\overline{\varphi}(x) + \theta\xi(x;t_0)\overline{\varphi}(x))\,dx \sim \frac{1}{2}\theta^2 \sup_{t\ge 0}\int\limits_0^\infty \xi(x;t_0)^2\overline{\varphi}(x)\,dx.$$

Тогда с учетом того, что $h(x) = \xi(x;t_0)\overline{\varphi}(x)$, получаем

$$\begin{aligned} e_B &= \lim_{\theta \to 0} \frac{(b_D(\theta))^2}{8(\delta^2(t_0))K(\theta)} \\ &= \left( \int\limits_0^\infty \xi(x;t_0)h(x)\,dx \right)^2 \Big/ \left( \delta(t)^2(t_0) \int\limits_0^\infty \xi(x;t_0)^2 \overline{\varphi}(x)\,dx \right) \\ &= \frac{(\delta^2)^2(t_0)}{(\delta^4(t_0))} = 1. \end{aligned}$$

Таким образом, при альтернативной плотности (19), достигается локальная асимптотическая эффективность.

## 7. Заключение

В данной дипломной работе были рассмотрены два критерия для проверки нормальности, основанные на свойстве нормального закона. Были исследованы предельные распределения и большие уклонения построенных статистик интегрального типа и типа Колмогорова, а также вычислена локальная эффективность по Бахадуру для ряда альтернатив. Для многих из рассмотренных альтернатив эффективность оказалась равна нулю. Тем не менее,как показано в работе, для обоих типов статистик существуют альтернативы специального вида, для которых рассматриваемые статистики локально асимптотически оптимальны по Бахадуру.

# Список литературы


1. Ahsanullah M., Kibria B.G., Shakil M. *Normal and Student's t-distributions and their applications.* Springer Science & Business Media, (2014)
2. Кобзарь, *А.И. Прикладная математическая статистика.* М:Физматлит, (2006)
3. Каган, А.М., Линник, Ю.В.,Рао С.Р. *Характеризационные задачи математической статистики.* М., Наука, (1973)
4. Arnold B. C., Gomez H. W., Salinas H. S. *A doubly skewed normal distribution, Statistics.* A Journal of Theoretical and Applied Statistics, (2014)
5. Bryc W. *The normal distribution: characterizations with applications.* Springer Science & Business Media, (2012) – V. 100.
6. Gradshteyn, I. S.; Ryzhik, I.M.; Geronimus, Y. V.; Tseytlin, M.Y. Jeffrey, Alan; Zwillinger, Daniel, eds. *Table of Integrals, Series, and Products.* (2007)
7. Hoeffding W. *A Class of Statistics with Asymptotically Normal Distribution.* Ann. Math. Statist., (1948), 293-325.
8. Polya G. *Herleitung des Gauss'chen Fehlergesetzes aus einer Funktionalgleichung.* Mathematische Zeitschrift, 18(1923), N 1, 96-108.
9. Marsaglia G . *Ratios of Normal Variables and Ratios of Sums of Uniform Variables.* Journal of the American Statistical Association, 60(1965), 193–204.
10. Muliere, P., Nikitin, Y.Y. *Scale-invariant test of normality based on Polya's characterization.* Metron, 60(2002), N 1-2, 21 - 33.
11. Litvinova V., Nikitin Y. *Kolmogorov Tests of Normality Based on Some Variants of Polya's Characterization.* Journal of Mathematical Sciences, 219(2016), 782-788.
12. Litvinova V. V., Nikitin Y. Y. *Two families of normality tests based on Polya-type characterization and their efficiencies.* Journal of Mathematical Sciences. 139(2006), N 3, 6582-6588.
13. Nikitin, Y.Y. *Tests based on characterizations, and their efficiencies: a survey.* Acta et Commentationes Universitatis Tartuensis de Mathematica, 21(2017), N 1, 3-24.
14. Nikitin Y. Y., Volkova K. Y. *Efficiency of exponentiality tests based on a special property of exponential distribution.* Mathematical Methods of Statistics, 25(2016), N 1, 54-66.
15. Mauldon J. G. *Characterizing properties of statistical distributions.* The Quarterly Journal of Mathematics, 7(1956), N 1, 155-160.

16. Azzalini A., *The Skew-Normal and Related Families.* Cambridge University Press, (2014).
17. Mudholkar, Govind, Hutson, Alan. *The epsilon-skew-normal distribution for analyzing near-normal data.* Journal of Statistical Planning and Inference, 83(2000), 291-309.
18. Ng, E. W., Geller, M. *A table of integrals of the error functions.* Journal of Research of the National Bureau of Standards, 73B(1969), 1-20.
19. Owen, D. *A table of normal integrals.* Communications in Statistics: Simulation and Computation, B9(1980), 389–419.
20. Laha R. G. *An example of a nonnormal distribution where the quotient follows the Cauchy law.* Proceed. of the Nation. Acad. of Sciences, 44(1958), N 2, 222-223.
21. Kotlarski I. *On bivariate random variables where the quotient of their coordinates follows some known distribution.* Annals of Mathem. Statist, 35(1964), N 4, 1673-1684.
22. Nikitin Ya. Yu., Peaucelle I. *Efficiency and Local Optimality of Distribution-Free Tests Based on $U$- and $V$-Statistics.* Metron, LXII(2004), 185-200.
23. Королюк В. С., Боровских Ю. В. *Теория $U$-статистик.* Наукова Думка, (1989).
24. Bahadur, R.R. *Some limit theorems in statistics.* SIAM:Philadelphia, (1971).
25. Никитин, Я.Ю. *Асимптотическая эффективность непараметрических критериев.* М.: Наука, (1995).
26. Никитин Я. Ю., Поникаров Е. В. *Грубая асимптотика вероятностей больших уклонений черновского типа для функционалов Мизеса и $U-$статистик.* Труды Санкт-Петербургского математического общества, 7(1999), 23–47.
27. Nikitin Ya. Yu. *Large deviations of $U$-empirical Kolmogorov-Smirnov tests, and their efficiency.* J. Nonpar. Stat., 22(2010), 649–668.